Wheaton College (MA)

COMP-401

Senior Seminar

AGI and the Near Future

Author:
Bryan Jensen

Professor:
Tom Armstrong

March 17, 2015

Contents

1   Introduction

Artificial Intelligence (AI) has been in the news quite a lot lately, and in a different light than ever before. Big names in technology and science have been voicing their concerns in interviews and over social media about possible imminent dangers of AI. The people voicing these concerns - public figures in science and technology such as Elon Musk[11] and Stephen Hawking[8] - are not directly related to the research of AI, but are generally considered among the most intelligent of our time.

Their concerns are related to the possibility of developing an artificial intelligence that far exceeds human intelligence, and about the kind of possibly uncontrollable power that would entail. Since an AI along these lines could be used to improve itself, we would likely notice a rapidly increasing feedback loop of progress, building towards an event often deemed "The Singularity"[14]. This amount of power could rival the nuclear bomb, and since it is itself intelligent, we would have no guarantee of our ability to control it.

The main impetus for the recent levels of concern is a slew of new estimates for the proximity of the singularity. Experts have been optimistically quoted saying it will take place within the next few decades, possibly as soon as the year 2030[14]. Even the more skeptical agree that it will certainly happen within the century[14]. Just a quick note on the terminology: the term *singularity* is, in general, "a point at which a function takes an infinite value." Most commonly used for the density of matter at the center of a black hole, the *technological* singularity is used to refer to the moment when progress starts to accelerate to the extent where, to humans, it may as well be infinitely fast.

Part of what will feed into AI's ability to produce this exponential growth is the observed exponential growth in computing technology that has taken place over the past few decades. This is most often quoted as "Moore's Law," originally stated in 1965, which is quoted as saying that the density of transistors on consumer-grade processors would double roughly every two years. Note, the original Moore's Law time frame was 24 months, and lately - since 2013[3] - seems to be increasing to beyond 36. Even so, it is a large increase in computing power happening every year, and it could still allow for huge advances in what's technically possible.

2 Artificial Intelligence

2.1 AI in the "Real World"

AI is a popular concept in sci-fi books, movies and television, with entries such as iRobot and Her, boasting very idealistic representations such as humanoid robots and entities at least as smart as humans. However, real-world AI research currently trends more towards finding a solution to a very specific problem, using the simplification of that problem rather a generalization of the technique. What this most often means is that they take a problem that is viewed as "hard," such as the game of chess, and break it down into solvable, computable chunks. Often this ends up reducing the given problem into a tree, graph or other searchable space, against which we have many well developed tools and algorithms we can leverage.

Another form is a rule-based structure that matches situations with solutions for that situation. This is popular in video gaming, where events can trigger reactions in the AI, intended to address that event to the furthering the AI's goals. For example, when presented with a human player, a bot may employ a tactic to play aggressively and kill the human player, thus completing the corresponding goal pre-programmed into the "bot."

2.2 What *I* Consider Artificial Intelligence

My personal definition of AI stresses the *intelligence* half of the concept, defining it as a process that learns - this idea happens to encompass a suite of algorithms referred to as "genetic algorithms." The central concept behind these algorithms is to mimic biological evolution and thereby learn how to solve a problem starting from complete ignorance and eventually, in theory, becoming a master. The key to these algorithms is that they themselves learn how to solve the problem, and do not rely any outside knowledge.

2.3 Real Examples of AI

There are numerous examples of AI in the world at the moment. About two decades ago, the most popular of these was DeepBlue, the computer built solely to win chess against the world grand master[2]. Another famous example is Watson,

designed by IBM to be capable of beating the world's best players at the TV Show Game *Jeopardy!*[9], which it managed to do handily to a large audience[7]. A final example, one less well known, is an attempt by Google to design a system to recognize text in pictures, a system which was actually capable of breaking the spambot-filter systems throughout the internet known as the "captcha" or "reCaptcha". These are all examples of "Artificial Narrow Intelligence," the most common type of AI in the world today.

3 Progression of AI

3.1 Artificial Narrow Intelligence

ANI covers everything in the world of practical AI in the world today. Also called *Weak AI*[5], it does one job, perhaps exceptionally well, but it does nothing else. For example, Watson is an impressive feat of programming and technology, but it only does what it does. Any attempt to use it to play a game of chess, and it will have no idea what to do.

3.2 Artificial General Intelligence

AGI, often called *Strong AI*[1] covers the entire concept of a form of AI that can compete with a human in every respect, from chess to talking to figuring out how to make a cup of coffee. This has been the goal of AI research since its inception, and is expected to be the invention that sparks the singularity.

3.3 Artificial Super Intelligence

An AI in this category can conceive of concepts incomprehensible to a human[14]. An ASI entity trying to explain what it knows to us would be analogous to humans trying to explain Einstein's relativity to a chimpanzee; there is simply no understanding on the behalf of the recipient. It is expected that once we achieve AGI, that entity's abilities can be used to improve itself, repeatedly, creating a feedback loop that would quickly bring us to an ASI.

With how quickly that can escalate, it is no surprise that the proximity of AGI is what gets the majority of the focus when talking about advances in technology

and artificial intelligence.

4   Moore's Law

However, the vast majority of talk and predictions about the upcoming singularity base its inevitability on the steady and massive increases in technological power at our disposal[13]. The thinking is, if we develop a powerful enough computer, then our algorithm doesn't have to be efficient to be smart, and by increasing that power we can eventually develop a system smarter than humans in every way.

The problem with this scenario is its reliance on Moore's Law. As stated before, Moore's Law has held true for half a century of computational advances, so why would it come into question now? The issue is that, while it has held for this long, it has slowed over time and there is no reason to believe it will alone provide the driving power for the creation of AGI.

4.1   What is Moore's Law?

Moore's Law, at its essence, dictates that computing power increases at an exponential rate with regard to time. Its officially endorsed form states that "the number of transistors per square inch on consumer-grade processors will double roughly every two years."[4]

4.2   What's Wrong With Moore's Law?

The law, stated in 1965, has largely retained its truthfulness despite the tests of time. Partially this is because it quickly became a self-fulfilling prophecy; following its inception, it soon became the industry goal, and hence spawned a second, sister Moore's Law (also known as Rock's Law): the cost of an Research & Development lab would also increase exponentially over time. With money being a finite resource, it is understandable why both of these could not continue forever[12].

However, Moore's Law *has* encountered and overcome hurdles in the past[10][6]. I don't have the space to enumerate them in this paper, but one such is the processor paradigm shift from single-core to multi-core, around 2006. So despite

the nay-sayers, Moore's law seems to have endured the tests of time, at least thus far.

4.3 "AGI can pick up where Moore's Law drops off"

Assuming we can continue the march of Moore's Law long enough to develop full AGI, the common thinking is that the new entity will be smart enough to help develop its own advances in computing power[14], and solve that issue while simultaneously improving itself in other ways. There is utter faith that if we manage to develop full ASI it will be able to solve all such problems, as its cognitive abilities will be just that far beyond ours - we may not understand it, but we believe in it. Therefore, Moore's Law does not have to persist forever, but just until we achieve a fully functional AGI.
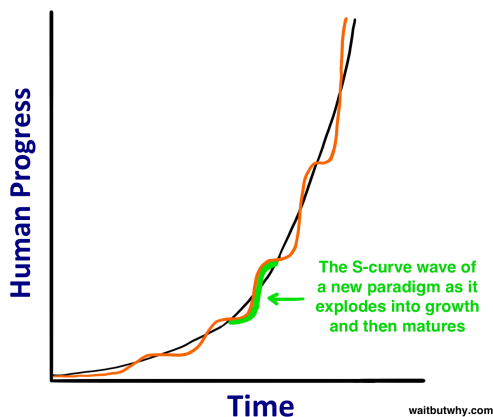
5 Inevitable Progress

Figure 1: Inexorable Progress

Even without Moore's Law, there is one fact that many quote as the driving reason why we must reach AGI (and, soon after, ASI): human progress is, historically, exponential[14]. It may seem to slow for periods at a time, but eventually we make such leaps and bounds that society quickly becomes unrecognizable. Society is vastly different in 2015 from that of 1800, due to the progress we've made in the past 200 years. But we would have much further back, much more than 200 years, from 1800, in order to reach the same level of change.

However, despite our newfound assurance that progress *will* happen, we still seem to be incapable of predicting what form it will take. It has been decades since we were first convinced that flying cars were to be the way of the future, and

as of 2015 we still drive ours along the same old paved roads.

Hence my argument becomes not that the human race *won't* progress, but rather that we simply don't know what that innovation will look like. For certain we cannot know that ASI is the form our next big leap will take.

6   Conclusion

I will not say that AGI and ASI are impossible parts of our (very near) future. Rather, I wish to emphasize that, with exponential growth meaning quicker and quicker change, we are in a time where we are suddenly unable to predict even a decade into the future; those ten years may result in more change than the entire past century. However, it is also possible that 2025 may come and go without much change from our current situation.

I will not, however, believe claims stating that AGI is imminent purely based on the inexorable march of Moore's Law. While Moore's Law is projected to hold out for at least another decade, even today its progress is slowing and it cannot be relied upon as the driving force behind producing *true* artificial intelligence. What I will believe is, within my lifetime, we will see many more changes that dramatically alter society, quite likely more drastically than in the past *millennium.* But we have no way of knowing what form those changes will take, or even if they will be instigated by technology as we currently understand it.

Artificial Narrow Intelligence is a powerful force in our world today. Artificial General Intelligence has the power to easily change our world. Artificial Super Intelligence has the power to easily *end* our world. But none of it is guaranteed to happen, at least not as the experts seem convinced is imminent.

References

[1] Artificial general intelligence. `http://en.wikipedia.org/wiki/Artificial_general_intelligence`. Accessed: 2015-03-12.

[2] Deep blue (chess computer). `http://en.wikipedia.org/wiki/Deep_Blue_%28chess_computer%29`. Accessed: 2015-03-12.

[3] International technology roadmap for semiconductors. `http://en.wikipedia.org/wiki/International_Technology_Roadmap_for_Semiconductors`. Accessed: 2015-03-12.

[4] Moore's law. `http://en.wikipedia.org/wiki/Moore%27s_law`. Accessed: 2015-03-15.

[5] Weak ai. `http://en.wikipedia.org/wiki/Weak_AI`. Accessed: 2015-03-12.

[6] ZiNgA BuRgA. Intel cpu speeds over time. `http://i.imgur.com/NFysh.png`. Accessed: 2015-03-03.

[7] Adam Gabbatt. Ibm computer watson wins jeopardy clash. `http://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy`. Accessed: 2015-03-12.

[8] Stephen Hawking. Stephen hawking warns artificial intelligence could end mankind. `http://www.bbc.com/news/technology-30290540`. Accessed: 2015-03-12.

[9] IBM. Ibm watson. `http://www.ibm.com/smarterplanet/us/en/ibmwatson/`. Accessed: 2015-03-12.

[10] John Markoff. I.b.m. researchers find a way to keep moore's law on pace. `http://www.nytimes.com/2006/02/20/technology/20chip.html?_r=0`. Accessed: 2015-03-03.

[11] Elon Musk. One-on-one with elon musk. `http://webcast.amps.ms.mit.edu/fall2014/AeroAstro/index-Fri-PM.html`. Accessed: 2015-03-12.

[12] Gabriel Popkin. Moores law is about to get weird. `http://nautil.us/issue/21/information/moores-law-is-about-to-get-weird`. Accessed: 2015-03-03.

[13] Tim Urban. The ai revolution: Our immortality or extinction. `http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html`. Accessed: 2015-03-12.

[14] Tim Urban. The ai revolution: The road to superintelligence. `http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html`. Accessed: 2015-03-03.